様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２９　年　　６　月　１６　日現在

機関番号：　１３９０１
研究種目：　基盤研究(B)（一般）
研究期間：　2014～2016
課題番号：　２６２８００６０

研究課題名（和文）統計的手法と生成過程モデリング手法の融合に基づく音声生成機能拡張技術の構築

研究課題名（英文）Development of augmented speech production techniques based on combination of statistical approaches and speech production modeling approaches

研究代表者

戸田　智基（Toda, Tomoki）

名古屋大学・情報基盤センター・教授

研究者番号：９０４０３３２８

交付決定額（研究期間全体）：（直接経費）　　12,500,000円

研究成果の概要（和文）：本研究では、音声生成過程における物理的制約により生じる障壁を取り除くために、音声生成機能を拡張するための基盤技術および応用技術の研究開発に取り組んだ。高精度な音声変換処理を可能とする統計的手法と、発声器官動作操作による音声変換処理を可能とする音声生成過程モデリング手法を融合させることで、現存の音声生成過程との親和性に優れ、かつ、高品質な変換処理を実現する音声変換手法を構築した。また、音声生成機能拡張の応用技術として、失われた声を再び取り戻すための発声障碍者補助技術、個人性を保持した外国語発声の生成技術、体内伝導音声を活用した周囲環境に頑健な通話技術などを構築した。

研究成果の概要（英文）：In this research, we developed fundamental techniques for augmented speech production and its applications to break down existing barriers caused by physical constraints in our speech production. High-quality speech conversion methods to be effectively used in our physical speech production mechanism were successfully developed by combining a statistical approach capable of generating high-quality converted speech and a speech production modeling approach capable of intuitively controlling converted speech by manipulating movements of speech organs. Moreover, we developed various applications of the augmented speech production techniques, such as a speaking aid technique towards restoration of lost voices, a foreign speech generation technique while keeping speaker identity, and a telecommunication technique using body-conducted speech.

研究分野：　音メディア情報処理

キーワード：　音声変換　音声合成　信号処理　統計処理　機能拡張

１．研究開始当初の背景

(1) 我々は個々の発声器官を巧みに操ることで、所望の音声信号を生成する。その際、生成される音声は、音声生成過程における物理的制約の影響を強く受ける。例えば、一か所でも発声器官が正常に動作しなくなると、深刻な発声障害を患う。このような障壁を無くすためには、物理的制約を超えた音声生成機能の拡張が必要となる。

(2) 既存の音声生成機能に対し、音声変換技術を組み合わせることで、仮想的に音声生成機能を拡張する技術の研究が行われている。
① 核となる技術として、学習データを用いて所望の音声変換処理を実現する統計的音声変換技術が研究されている。機械学習の発展に伴い、性能は着実に改善されているものの、処理が完全にブラックボックス化されるため、実現可能な変換処理は、システム学習時に用いる音声データに完全に依存する。そのため、人が自身の発声器官を巧みに動かして所望の音声を生成するように、直感的に変換音声を制御・調整することは難しい。
② 一方で、音声生成過程を数理的にモデル化する技術も古くから研究が進められている。音声生成過程モデリング技術では、発声器官動作と音声信号を明示的に結びつけることができるため、発声器官動作操作による音声変換処理を実現できる。しかし、モデリングの際に強い近似を必要とし、結果として生成される音声の品質劣化は避けられない。近似の少ない精密な物理モデルも研究されているが、計算量が爆発的に増加するため、即時性が重要となる音声生成機能拡張においては、その効果を十分に発揮できない。

２．研究の目的

(1) 統計的手法と生成過程モデリング手法は、各々異なる利点、欠点を持ち、それらは相補的な関係にある。そこで、統計的手法と生成過程モデリング手法を融合することで、発声器官動作操作機能を備えた音声変換基礎技術を構築し、音声生成機能を拡張する応用技術へと発展させることを目的とした。

(2) 基礎技術として、発声器官動作と音声信号間を対応付ける生成過程モデル、および、器官動作制御モデルを緩やかな制約条件として、音声信号の確率的生成モデルを定式化することで、統計的音声変換手法との融合を図った。また、そのような制約条件を得るために、調音動作と音声信号の同期収録データベースの構築を目的とした。さらに、
① 失われた声の回復を目指した発声障碍者補助技術、
② 声質および韻律的な個人性を制御可能とする外国語発声の生成技術、
③ 体内伝導音声を活用したサイレント通話技術および雑音下での音声強調技術、
といった応用技術の構築を目的とした。

３．研究の方法

(1) 調音動作操作機能を備えた統計的声質変換技術の構築：調音動作と音声信号間の対応関係の統計的モデリング技術を発展させ、音声信号からの調音動作推定、調音動作からの音声信号生成、調音動作操作による生成音声の制御・調整を可能とする統計的声質変換技術の構築に取り組んだ。

(2) 音源生成器官動作操作機能を備えた統計的韻律変換技術の構築：基本周波数（F_0）生成過程を考慮した確率モデルを発展させ、F_0 パターンからの音源生成器官動作指令推定、動作指令からの F_0 パターン生成、動作指令操作による F_0 パターン制御・調整を可能とする統計的韻律変換技術の構築に取り組んだ。

(3) 両技術を統合した統計的音声変換技術の構築：統計的声質変換処理と統計的韻律変換処理を統合し、調音動作と音源生成器官動作指令の操作機能を備えた統計的音声変換技術の構築に取り組んだ。

(4) 調音動作・音声同期収録データベースの構築：時間分解能に優れた磁気センサシステム（Electromagnetic articulography: EMA）を用いて、調音動作・音声同期収録を実施し、データベースの構築に取り組んだ。

(5) 音声生成機能を拡張する複数の応用技術の構築：音声生成機能拡張の応用技術として、以下の３つの技術構築に取り組んだ。
① 発声障害者補助技術：喉頭摘出者を対象とした統計的無喉頭音声強調技術を基盤技術とし、統計的韻律変換技術を導入することで、音源生成過程を考慮した F_0 パターンの予測処理の実現に取り組んだ。
② 個人性を保持した外国語発声の生成技術：外国語発声に対する統計的声質変換技術を基盤技術とし、異なる言語間にわたり観測される個人性に寄与する特徴量の分析を実施した。また、分析結果に基づき、外国語発声の声質・韻律に対する個人性制御技術の構築に取り組んだ。
③ 体内伝導音声を用いた音声強調技術：体表密着型マイクロフォンを用いて収録される体内伝導音声を用いて、周囲に聞こえないぐらい小さな音声を強調する技術と、騒音下で収録された音声を強調する技術の構築に取り組んだ。

４．研究成果

(1) 調音動作操作機能を備えた統計的声質変換技術の構築：統計的声質変換の基礎技術を拡張し、音声信号からの調音動作パラメータ推定処理、および、調音動作パラメータからの音声信号生成処理を実現し、これらの処理を繋ぎ合わせることで、調音動作操作機能を備えた統計的声質変換技術を構築した。ま

た、調音動作操作時に適切な調音動作を保証するために、調音動作パラメータ補正技術を構築した。さらに、品質劣化を生み出す主要因であるボコーダによる波形合成処理を回避するために，入力音声波形の直接加工処理に基づく統計的声質変換技術を構築した。本手法を，調音動作操作機能を備えた統計的声質変換技術へと適用し，高い音質を保持したまま調音動作操作による音声変換処理が可能であることを示した。

(2) 音源生成器官動作操作機能を備えた統計的韻律変換技術の構築：*F*0 パターン生成過程の確率モデルを考案し、音声信号からの音源生成器官動作指令推定技術を構築し、動作指令操作による *F*0 パターン変換処理を実現した。また、言語情報からの *F*0 パターン生成技術も構築した。さらに、統計的 *F*0 パターン予測モデルと *F*0 パターン生成過程モデルを、Product-of-Experts の枠組みで確率的に統合する手法を提案し，物理的制約を満たす統計的 *F*0 パターン予測処理を可能とした。

(3) 両技術を統合した統計的音声変換技術の構築：調音動作操作機能を備えた統計的声質変換技術と、音源生成器官動作操作機能を備えた統計的韻律変換技術を併用することで、発声器官動作制御機能を備えた統計的音声変換技術を構築した。さらに、音声波形加工に基づく韻律変換技術や、高精度な音声特徴量時系列モデリング技術の構築に取り組み、変換性能のさらなる改善に成功した。

(4) 調音動作・音声同期収録データベースの構築：EMA に基づくリアルタイム発話観測システムを用いて、調音動作と音声信号の同期収録を実施した。男性話者 8 名および女性話者 1 名の計 9 名を対象とし、各話者 50～100 文程度収録した。また、データベースの構築に向けて、データ整備に取り組んだ。

(5) 音声生成機能を拡張する複数の応用技術の構築：統計的手法に基づく実時間声質変換基盤技術の性能を改善し、音声生成機能拡張技術として、発声障碍者補助技術、外国語発声生成技術、体内伝導音声強調技術、さらには、ボイスチェンジャ／ボーカルエフェクタ技術を構築した。

① 発声障害者補助技術：物理的制約を考慮した統計的 *F*0 パターン予測技術を喉頭摘出者の電気音声強調に適用することで、従来の発声補助器具と比較して、より自然な音声を生成できることを示した。また、短遅延予測処理手法を提案し、その有効性を示した。

② 個人性を保持した外国語発声の生成技術：調音動作パラメータ操作に基づく発音補正技術を構築し、高音質な補正処理が可能であることを示した。また、日本語話者の英語音声に対して、個人性を保持しつつ自然性を改善する技術を構築し、その有効性を示した。

③ 体内伝導音声を用いた音声強調技術：体内伝導マイクロフォンと空気伝導マイクロフォンの併用により、体内伝導音声収録における外部雑音の影響を大幅に緩和する技術を構築し、その有効性を示した。また、さらに改良を加え、統計的体内伝導音声強調に適した外部雑音除去技術へと発展させた。

④ ボイスチェンジャ／ボーカルエフェクタ技術：知覚尺度に基づき声質を制御する技術を考案し、その有効性を示した。また、入力音声波形加工技術を導入することで、特定話者／歌手の声へと、リアルタイムかつ高音質に変換する音声変換システムの開発に成功した。なお、本システムは、国際的評価会 Voice Conversion Challenge 2016 (VCC2016) において、参加 17 機関中、最高性能の評価を得るに至った（図 1 参照）。また、企業との共同研究を通して、高品質なリアルタイム統計的ボイスチェンジャを実用化した。

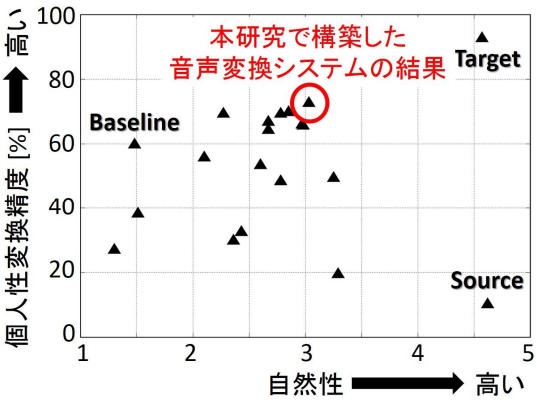

図 1　VCC2016 の評価結果

(6) これらの研究成果をとりまとめ、国内外において多数の研究発表を行った。本研究成果は高い評価を受け、国内外において計 8 つの賞を受賞し、また、10 件の招待講演を実施するに至った。

5．主な発表論文等

〔雑誌論文〕（計 43 件）

① Kou Tanaka、Tomoki Toda、Satoshi Nakamura、A vibration control method of an electrolarynx based on statistical F0 pattern prediction、IEICE Transactions on Information and Systems、査読有、Vol. E100-D、No. 9、2017、印刷中.

② Yuji Oshima、Shinnosuke Takamichi、Tomoki Toda、Graham Neubig、Sakriani Sakti、Satoshi Nakamura、Non-native text-to-speech preserving speaker individuality based on partial correction of prosodic and phonetic characteristics、IEICE Transactions on Information and Systems、査読有、Vol. E99-D、No. 12、2016、pp. 3132－3139.

DOI: 10.1587/transinf.2016EDP7231

③ Kazuhiro Kobayashi、Tomoki Toda、Tomoyasu Nakano、Masataka Goto、Satoshi Nakamura、Improvements of voice timbre control based on perceived age in singing voice conversion, IEICE Transactions on Information and Systems、査読有、Vol. E99-D、No. 11、2016、pp. 2767－2777.
DOI: 10.1587/transinf.2016EDP7234

④ Shinnosuke Takamichi、Tomoki Toda、Graham Neubig、Sakriani Sakti、Satoshi Nakamura、A statistical sample-based approach to GMM-based voice conversion using tied-covariance acoustic models、IEICE Transactions on Information and Systems、査読有、Vol. E99-D、No. 10、2016、pp. 2490－2498.
DOI: 10.1587/transinf.2016SLP0020

⑤ Tomoki Toda（1 番目）、他 6 名、The Voice Conversion Challenge 2016、Proceedings of INTERSPEECH、査読有、2016、pp. 1632－1636、2016.
DOI: 10.21437/Interspeech.2016-1066

⑥ 戸田 智基、はじめての音声変換、日本音響学会誌、査読無、Vol. 72、No. 6、2016、pp. 324－331、**招待形式解説論文**.

⑦ Shinnosuke Takamichi、Tomoki Toda、Alan W Black、Graham Neubig、Sakriani Sakti、Satoshi Nakamura、Post-filters to modify the modulation spectrum for statistical parametric speech synthesis、IEEE/ACM Transactions on Audio, Speech and Language Processing、査読有、Vol. 24、No. 4、2016、pp. 755－767.
DOI: 10.1109/TASLP.2016.2522655

⑧ Hirokazu Kameoka（1 番目）、他 5 名、Generative modeling of voice fundamental frequency contours、IEEE/ACM Transactions on Audio, Speech and Language Processing、査読有、Vol. 23、No. 6、2015、pp. 1042－1053.
DOI: 10.1109/TASLP.2015.2418576

⑨ Shinnosuke Takamichi、Tomoki Toda、Alan W. Black、Satoshi Nakamura、Modulation spectrum-based post-filter for GMM-based voice conversion、Proceedings of APSIPA ASC、査読有、2014、pp. 1－4、2014、**APSIPA ASC 2014 The Best Paper Award.**
10.1109/APSIPA.2014.7041540

⑩ Tomoki Toda、Augmented speech production based on real-time statistical voice conversion、Proceedings of GlobalSIP、査読有、2014、pp. 592－596、**招待講演論文**.
10.1109/GlobalSIP.2014.7032186

（この他に 33 件）

〔学会発表〕（計 76 件）

① 戸田 智基、音声信号の分析と加工 – 音声を自在に変換するには？、日本音響学会春季研究発表会、2017 年 3 月 15 日、明治大学（神奈川県川崎市）、**招待講演**.

② 田尻 祐介、亀岡 弘和、戸田 智基、実環境下におけるサイレント音声通話の実現に向けた雑音環境変動に頑健な非可聴つぶやき強調法、第 3 回サイレント音声認識ワークショップ、2016 年 10 月 14 日、福岡朝日ビル（福岡県福岡市）、**学生奨励賞**.

③ 亀岡 弘和、統計的音響信号処理、NLP 若手の会(YANS)第 11 回シンポジウム、2016 年 8 月 28 日、ホテルシーモア（和歌山県西牟婁郡）、**招待講演**.

④ 戸田 智基、音情報処理における特徴表現、第 19 回画像の認識・理解シンポジウム（MIRU2016）、特別企画 MIRU x KIKU、2016 年 8 月 3 日、アクトシティ浜松（静岡県浜松市）、**招待講演**.

⑤ 亀岡 弘和、音響信号の分解と再構成、第 19 回画像の認識・理解シンポジウム（MIRU2016）、特別企画 MIRU x KIKU、2016 年 8 月 3 日、アクトシティ浜松（静岡県浜松市）、**招待講演**.

⑥ 田尻 祐介、亀岡 弘和、戸田 智基、中村 哲、空気／体内伝導信号の非負値テンソル分解に基づく体内伝導微弱音声に対する雑音抑圧法、日本音響学会春季研究発表会、2016 年 3 月 9 日、桐蔭横浜大学（神奈川県横浜市）、**第 13 回日本音響学会学生優秀発表賞**（受賞者：田尻祐介）.

⑦ 田中 宏、戸田 智基、Graham Neubig、Sakriani Sakti、中村 哲、統計的手法を用いた電気式人工喉頭制御における遅延時間と予測精度の調査、日本音響学会秋季研究発表会、2015 年 9 月 18 日、会津大学（福島県会津若松市）、**第 12 回日本音響学会学生優秀発表賞**（受賞者：田中宏）.

⑧ 高道 慎之介、戸田 智基、Alan W Black、中村 哲、統計的パラメトリック音声合成のための変調スペクトルに基づく音質改善法、情報処理学会音楽情報科学研究会、2015 年 5 月 24 日、電気通信大学（東京都調布市）、**2015 年度音楽情報科学研究会学生賞**（受賞者：高道 慎之介）.

⑨ Patrick Lumban Tobing、Kazuhiro Kobayashi、Tomoki Toda、Graham Neubig、Sakriani Sakti、Satoshi Nakamura、Articulatory controllable speech modification based on gaussian mixture models with direct waveform modification using spectrum differential、日本音響学会春季研究発表会、2015 年 3 月 17 日、中央大学後楽園キャンパス（東京都文京区）、第 11 回**日本音響学会学生優秀発表賞**（受賞者：Patrick Lumban Tobing）.

⑩ 高道 慎之介、戸田 智基、Alan W Black、中村 哲、統計的パラメトリック音声合成のための変調スペクトル制約付きトラジェクトリ学習アルゴリズム、電子情報通信学会／日本音響学会 音声研究会、2015年3月2日、南の美ら花ホテルミヤヒラ（沖縄県石垣市）、**2014年度 音声研究会研究奨励賞**（受賞者：高道 慎之介）.

⑪ 小林 和弘、戸田 智基、中野 倫靖、後藤真孝、Graham Neubig、Sakriani Sakti、中村 哲、性別依存重回帰混合正規分布モデル基づく差分スペクトル補正による歌声の知覚年齢制御法、日本音響学会秋季研究発表会、2014年9月5日、北海学園大学豊平キャンパス（北海道札幌市）、**第10回日本音響学会学生優秀発表賞**（受賞者：小林 和弘）.

（この他に65件）

〔図書〕（計3件）

① Hirokazu Kameoka、Springer Japan、Applied Matrix and Tensor Variate Data Analysis、分担「Non-negative matrix factorization and its variants for audio signal processing」、2016、pp. 23－50.

② 戸田 智基、コロナ社、音響キーワードブック、分担「声質変換」、2016、pp. 260－261.

③ 戸田 智基、近代科学社、シンギュラリティ 限界突破を目指した最先端研究、分担「声とその表情を生み出すコンピュータ」、2016、pp. 176－181.

〔産業財産権〕

○出願状況（計6件）

名称：信号解析装置、方法、及びプログラム
発明者：亀岡 弘和、田尻 祐介、戸田 智基、中村 哲
権利者：日本電信電話株式会社、国立大学法人奈良先端科学技術大学院大学
種類：特許
番号：特許願2016－032414
出願年月日：平成２８年2月２３日
国内外の別：国内

名称：基本周波数パターン予測装置、方法、及びプログラム
発明者：亀岡 弘和、田中 宏、戸田 智基、中村 哲
権利者：日本電信電話株式会社、国立大学法人奈良先端科学技術大学院大学
種類：特許
番号：特許願2016－032411
出願年月日：平成２８年2月２３日
国内外の別：国内

名称：電気式人工喉頭装置
発明者：戸田 智基、田中 宏、中村 哲、サクティ サクリアニ、ニュービッグ グラム
権利者：国立大学法人奈良先端科学技術大学院大学
種類：特許
番号：特許願2015－530782
出願年月日：平成２８年1月２６日
国内外の別：国内

名称：肉伝導音集音システム及び肉伝導音集音システムの装着方法
発明者：田尻 祐介、戸田 智基、ニュービッグ グラム、サクティ サクリアニ、中村 哲
権利者：国立大学法人奈良先端科学技術大学院大学
種類：特許
番号：特許願2015－170837
出願年月日：平成２７年8月３１日
国内外の別：国内

（この他に2件）

６．研究組織

(1)研究代表者

戸田　智基（TODA, Tomoki）
名古屋大学・情報基盤センター・教授
研究者番号：９０４０３３２８

(2)研究分担者

亀岡　弘和（KAMEOKA, Hirokazu）
日本電信電話株式会社NTTコミュニケーション科学基礎研究所・メディア情報研究部・主任研究員／特別研究員
研究者番号：２０４６６４０２

猿渡　洋（SARUWATARI, Hiroshi）
東京大学・情報理工学（系）研究科・教授
研究者番号：３０３２４９７４

中村　哲（NAKAMURA, Satoshi）
奈良先端科学技術大学院大学・情報科学研究科・教授
研究者番号：３０２６３４２９

サクティ　サクリアニ（SAKTI, Sakriani）
奈良先端科学技術大学院大学・情報科学研究科・助教
研究者番号：３０６２５０８３
2016年3月まで研究分担者として参加

ニュービッグ　グラム（NEUBIG, Graham）
奈良先端科学技術大学院大学・情報科学研究科・助教
研究者番号：７０６３３４２８
2016年3月まで研究分担者として参加

川波　弘道（KAWANAMI, Hiromichi）
奈良先端科学技術大学院大学・情報科学研究科・助教
研究者番号：８０３３５４８９
2016年3月まで研究分担者として参加